05-24-00 A

Please type a plus sign (+) inside this box [+]

PTO/SB/05 (12/97)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

UTILITY PATENT APPLICATION TRANSMITTAL
(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No. 003932.P014 Total Pages 5

First Named Inventor or Application Identifier Matthew Lennig

Express Mail Label No. EL627466804US

ADDRESS TO: **Assistant Commissioner for Patents**
Box Patent Application
Washington, D. C. 20231

APPLICATION ELEMENTS
See MPEP chapter 600 concerning utility patent application contents.

1. X Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)

2. X Specification (Total Pages 28)
(preferred arrangement set forth below)
- Descriptive Title of the Invention
- Cross References to Related Applications
- Statement Regarding Fed sponsored R & D
- Reference to Microfiche Appendix
- Background of the Invention
- Brief Summary of the Invention
- Brief Description of the Drawings (if filed)
- Detailed Description
- Claims
- Abstract of the Disclosure

3. X Drawings(s) (35 USC 113) (Total Sheets 5)

4. X Oath or Declaration (Total Pages 5) **(UNSIGNED)**
   a. ___ Newly Executed (Original or Copy)
   b. ___ Copy from a Prior Application (37 CFR 1.63(d))
   (for Continuation/Divisional with Box 17 completed) **(Note Box 5 below)**
   i. ___ DELETIONS OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).

5. ____ Incorporation By Reference (useable if Box 4b is checked)
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

6. ____ Microfiche Computer Program (Appendix)

7. ______ Nucleotide and/or Amino Acid Sequence Submission
(if applicable, all necessary)
a. ____ Computer Readable Copy
b. ____ Paper Copy (identical to computer copy)
c. ____ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ______ Assignment Papers (cover sheet & documents(s))
9. ______ a. 37 CFR 3.73(b) Statement (where there is an assignee)
______ b. Power of Attorney
10. ______ English Translation Document (if applicable)
11. _X_ a. Information Disclosure Statement (IDS)/PTO-1449
_X_ b. Copies of IDS Citations
12. ______ Preliminary Amendment
13. _X_ Return Receipt Postcard (MPEP 503) (Should be specifically itemized)
14. ______ a. Small Entity Statement(s)
______ b. Statement filed in prior application, Status still proper and desired
15. ______ Certified Copy of Priority Document(s) (if foreign priority is claimed)

16. _X_ Other: Express Mail Certificate of Mailing with copy of postcard (in duplicate)

17. **If a CONTINUING APPLICATION,** check appropriate box and supply the requisite information:
____ Continuation ____ Divisional ____ Continuation-in-part (CIP)
of prior application No: _

18. **Correspondence Address**
______ Customer Number or Bar Code Label ______________________
(Insert Customer No. or Attach Bar Code Label here)
or
_X_ Correspondence Address Below

NAME Jordan M. Becker (Reg. No. 39,602)
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP

ADDRESS 12400 Wilshire Boulevard
Seventh Floor

CITY Los Angeles STATE California ZIP CODE 90025-1026

Country U.S.A. TELEPHONE (408) 720-8300 FAX (408) 720-9397

EXPRESS MAIL CERTIFICATE OF MAILING

"Express Mail" mailing label number: EL627466804US
Date of Deposit: May 22, 2000
I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Assistant Commissioner for Patents, Washington, D. C. 20231

Mary E. Buggie
(Typed or printed name of person mailing paper or fee)
Mary E. Buggie
(Signature of person mailing paper or fee)
5/22/2000
(Date signed)

Serial/Patent No.: ______ Filing/Issue Date: ______
Client: Nuance Communications
Title: PROSODY BASED ENDPOINT DETECTION

BSTZ File No.: 003932.P014 Atty/Secty Initials: JMB/meb
Date Mailed: May 22, 2000 Docket Due Date: ****

The following has been received in the U.S. Patent & Trademark Office on the date stamped hereon:

- ☐ Amendment/Response (____ pgs.)
- ☐ Appeal Brief (____ pgs.) (in triplicate)
- ■ Application - Utility (28 pgs., with cover and abstract)
- ☐ Application - Rule 1.53(b) Continuation (____ pgs.)
- ☐ Application - Rule 1.53(b) Divisional (____ pgs.)
- ☐ Application - Rule 1.53(b) CIP (____ pgs.)
- ☐ Application - Rule 1.53(d) CPA Transmittal (____ pgs.)
- ☐ Application - Design (____ pgs.)
- ☐ Application - PCT (____ pgs.)
- ☐ Application - Provisional (____ pgs.)
- ☐ Assignment and Cover Sheet
- ■ Certificate of Mailing Express-duplicate
- ■ Declaration & POA (5 pgs.) (unsigned)
- ☐ Disclosure Docs & Orig & Copy of Inventor's Signed Letter (____ pgs)
- ■ Drawings: 5 # of sheets includes 6 figures
- ■ Express Mail No.: EL627466804US
- ☐ ____ Month(s) Extension of Time
- ■ Information Disclosure Statement & PTO-1449 (4 pgs.)
- ☐ Issue Fee Transmittal
- ☐ Notice of Appeal
- ☐ Petition for Extension of Time
- ☐ Petition for ______
- ■ Postcard
- ☐ Power of Attorney (____ pgs.)
- ☐ Preliminary Amendment (____ pgs.)
- ☐ Reply Brief (____ pgs.)
- ☐ Response to Notice of Missing Parts
- ☐ Small Entity Declaration for Indep. Inventor/Small Business
- ■ Transmittal Letter, in duplicate
- ■ Fee Transmittal, in duplicate
- ■ Check No. 35455 Amt: $1,614.00
- ☐ Check No. ____ Amt: ____

■ Other: Two Cited References

UNITED STATES PATENT APPLICATION

FOR

Prosody Based Endpoint Detection

INVENTOR:

Matthew Lennig

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 WILSHIRE BOULEVARD
SEVENTH FLOOR
LOS ANGELES, CALIFORNIA 90025
(408) 720-8300

Attorney's Docket No. 003932.P014

"Express Mail" mailing label number EL627466804US

Date of Deposit May 22, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

Mary E. Buggie
(Typed or printed name of person mailing paper or fee)

Mary E. Buggie
(Signature of person mailing paper or fee)

Prosody Based Endpoint Detection

FIELD OF THE INVENTION

The present invention pertains to endpoint detection in the processing of speech, such as in speech recognition. More particularly, the present invention relates to the detection of the endpoint of an utterance using prosody.

BACKGROUND OF THE INVENTION

In a speech recognition system, a device commonly known as an "endpoint detector" separates the speech segment(s) of an utterance represented in an input signal from the non-speech segments, i.e., it identifies the "endpoints" of speech. An "endpoint" of speech can be either the beginning of speech after a period of non-speech or the ending of speech before a period of non-speech. An endpoint detector may be either hardware-based or software-based, or both. Because endpoint detection generally occurs early in the speech recognition process, the accuracy of the endpoint detector is crucial to the performance of the overall speech recognition system. Accurate endpoint detection will facilitate accurate recognition results, while poor endpoint detection will often cause poor recognition results.

Some conventional endpoint detectors operate using log energy and/or spectral information as knowledge sources. For example, by comparing the log energy of the input speech signal against a threshold energy level, an endpoint can be identified. An end-of-utterance can be identified, for example, if the log energy drops below the threshold level after having exceeded the threshold level for some specified length of time. However, this approach does not take into consideration many of the characteristics of human speech. As a result, this approach is only a rough approximation, such that purely energy-based endpoint detectors are not as accurate as desired.

One problem associated with endpoint detection is distinguishing between a mid-utterance pause and the end of an utterance. In making this determination, there is generally an inherent trade-off between achieving short latency and detecting the entire utterance.

SUMMARY OF THE INVENTION

A method and apparatus for performing endpoint detection are provided. In the method, a speech signal representing an utterance is input. The utterance has an intonation, based on which the endpoint of the utterance is identified. In particular embodiments, endpoint identification may include referencing the intonation of the utterance against an intonation model.

Other features of the present invention will be apparent from the accompanying drawings and from the detailed description which follows.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements and in which:

Figure 1 is a block diagram of a speech recognition system;

Figure 2 is a block diagram of a processing system that may be configured to perform speech recognition;

Figure 3 is a flow diagram showing an overall process for performing endpoint detection using prosody;

Figure 4 is a flow diagram showing in greater detail the process of Figure 3, according to one embodiment; and

Figures 5A and 5B are flow diagrams showing in greater detail the process of Figure 3, according to a second embodiment.

DETAILED DESCRIPTION

A method and apparatus for detecting endpoints of speech using prosody are described. Note that in this description, references to "one embodiment" or "an embodiment" mean that the feature being referred to is included in at least one embodiment of the present invention. Further, separate references to "one embodiment" in this description do not necessarily refer to the same embodiment; however, neither are such embodiments mutually exclusive, unless so stated and except as will be readily apparent to those skilled in the art.

As described in greater detail below, an end-of-utterance condition can be identified by an endpoint detector based, at least in part, on the prosody characteristics of the utterance. Other knowledge sources, such as log energy and/or spectral information may also be used in combination with prosody. Note that while endpoint detection generally involves identifying both beginning-of-utterance and end-of-utterance conditions (i.e., separating speech from non-speech), the techniques described herein are directed primarily toward identifying an end-of-utterance condition. Any conventional endpointing technique may be used to identify a beginning-of-utterance condition, which technique(s) need not be described herein. Nonetheless, it is contemplated that the prosody-based techniques described herein may be extended or modified to detect a beginning-of-utterance condition as well. The processes described herein are real-time processes that operate on a continuous audio signal, examining the incoming speech frame-by-frame to detect an end-of-utterance condition.

"Prosody" is defined herein to include characteristics such as intonation and syllable duration. Hence, an end-of-utterance condition may be identified based, at least in part, on the intonation of the utterance, the duration of one or more syllables of the utterance, or a combination of these and/or other variables. For example, in many languages, including English, the end of an utterance often has a generally decreasing intonation. This fact can be used to advantage in endpoint detection, as further described below. Various types of prosody models may be used in this process. This prosody based approach, therefore, makes use of more of the inherent features of human speech than purely energy-based

approaches and other more traditional approaches. Among other advantages, the use of intonation in the endpoint detection process helps to more accurately distinguish between a mid-utterance pause and an end-of-utterance condition, without adversely affecting latency. Consequently, the prosody based approach provides more accurate endpoint detection without adversely affecting latency and thereby facilitates improved speech recognition.

Figure 1 shows an example of a speech recognition system in which the present endpoint detection technique can be implemented. The illustrated system includes a dictionary 2, a set of acoustic models 4, and a grammar/language model 6. Each of these elements may be stored in one or more conventional storage devices. The dictionary 2 contains all of the words allowed by the speech application in which the system is used. The acoustic models 4 are statistical representations of all phonetic units and subunits of speech that may be found in a speech waveform. The grammar/language model 6 is a statistical or deterministic representation of all possible combinations of word sequences that are allowed by the speech application. The system further includes an audio front end 7 and a speech decoder 8. The audio front end includes an endpoint detector 5. The endpoint detector 8 has access to one or more prosody models 3-1 through 3-N, which are discussed further below.

An input speech signal is received by the audio front end 7 via a microphone, telephony interface, computer network interface, or any other suitable input interface. The audio front end 7 digitizes the speech waveform (if not already digitized), endpoints the speech (using the endpoint detector 5), and extracts feature vectors (also known as features, observations, parameter vectors, or frames) from the digitized speech. In some implementations, endpointing precedes feature extraction, while in other implementations feature extraction may precede endpointing. To facilitate description, the former case is assumed henceforth in this description.

Thus, the audio front end 7 is essentially responsible for processing the speech waveform and transforming it into a sequence of data points that can be better modeled by the acoustic models 4 than the raw waveform. The extracted

feature vectors are provided to the speech decoder 8, which references the feature vectors against the dictionary 2, the acoustic models 4, and the grammar/language model 6, to generate recognized speech data. The recognized speech data may further be provided to a natural language interpreter (not shown), which interprets the meaning of the recognized speech.

The prosody based endpoint detection technique is implemented within the endpoint detector 5 in the audio front end 7. Note that audio front ends which perform the above functions but without a prosody based endpoint detection technique are well known in the art. The prosody based endpoint detection technique may be implemented using software, hardware, or a combination of hardware and software. For example, the technique may be implemented by a microprocessor or Digital Signal Processor (DSP) executing sequences of software instructions. Alternatively, the technique may be implemented using only hardwired circuitry, or a combination of hardwired circuitry and executing software instructions. Such hardwired circuitry may include, for example, one or more microcontrollers, Application Specific Integrated Circuits (ASICs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), A/D converters, and/or other suitable components.

The system of Figure 1 may be implemented in a conventional processing system, such as a personal computer (PC), workstation, hand-held computer, Personal Digital Assistant (PDA), etc. Alternatively, the system may be distributed between two or more such processing systems, which may be connected on a network. Figure 2 is a high-level block diagram of an example of such a processing system. The processing system of Figure 2 includes a central processing unit (CPU) 10 (e.g., a microprocessor), random access memory (RAM) 11, read-only memory (ROM) 12, and a mass storage device 13, each connected to a bus system 9. Mass storage device 13 may include any suitable device for storing large volumes of data, such as magnetic disk or tape, magneto-optical (MO) storage device, or any of various types of Digital Versatile Disk (DVD) or compact disk (CD) based storage, flash memory, etc. The bus system 9 may include one or more buses connected to each other through various bridges,

controllers and/or adapters, such as are well-known in the art. For example, the bus system 9 may include a system bus that is connected through an adapter to one or more expansion buses, such as a Peripheral Component Interconnect (PCI) bus.

Also coupled to the bus system 9 are an audio interface 14, a display device 15, input devices 16 and 17, and a communication device 30. The audio interface 14 allows the computer system to receive an input audio signal that includes the speech signal. The audio interface 14 includes circuitry and (in some embodiments) software instructions for receiving an input audio signal which includes the speech signal, which may be received from a microphone, a telephone line, a network interface, etc., and for transferring such signal onto the bus system 9. Thus, prosody based endpoint detection as described herein may be performed within the audio interface 14. Alternatively, the endpoint detection may be performed within the CPU 10, or partly within the CPU 10 and partly within the audio interface 14. The audio interface may include one or more DSPs, general-purpose microprocessors, microcontrollers, ASICs, PLDs, FPGAs, A/D converters, and/or other suitable components.

The display device 15 may be any suitable device for displaying alphanumeric, graphical and/or video data to a user, such as a cathode ray tube (CRT), a liquid crystal display (LCD), or the like, and associated controllers. The input devices 16 and 17 may include, for example, a conventional pointing device, a keyboard, etc. The communication device 18 may be any device suitable for enabling the computer system to communicate data with another processing system over a network via a data link 20, such as a conventional telephone modem, a wireless modem, a cable modem, an Integrated Services Digital Network (ISDN) adapter, a Digital Subscriber Line (DSL) modem, an Ethernet adapter, or the like.

Note that some of these components may be omitted in certain embodiments, and certain embodiments may include additional or substitute components that are not mentioned here. Such variations will be readily apparent to those skilled in the art. As an example of such a variation, the functions of the

audio interface 14 and the communication device 18 may be provided in a single device. As another example, the peripheral components connected to the bus system 9 might further include audio speakers and associated adapter circuitry. As yet another example, the display device 15 may be omitted if the processing system has no direct interface to a user.

Prosody based endpoint detection may be based, at least in part, on the intonation of utterances. Of course, endpoint detection may also be based on other prosodic information and/or on non-prosodic information, such as log energy.

Figure 3 shows, at a high level, a process for detecting an end-of-utterance condition based on prosody, according to one embodiment. The next frame of speech representing at least part of an utterance is initially input to the endpoint detector 5 at 301. The end-of-utterance condition is identified at 302 based (at least) on the intonation of the utterance, and the routine then repeats. Note that this process and the processes described below are real-time processes that operate on a continuous audio signal, examining the incoming speech frame-by-frame to detect an end-of-utterance condition. For purposes of detecting an end-of-utterance condition, the time frame of this audio signal may be assumed to be after the start of speech.

As noted, other types of prosodic parameters and more traditional, non-prosodic knowledge sources can also be used to detect an end-of-utterance condition (although not so indicated in Figure 3). A technique for combining multiple knowledge sources to make a decision is described in U.S. Patent no. 5,097,509 of Lennig, issued on March 17, 1992 ("Lennig"), which is incorporated herein by reference. In accordance with the present invention, the technique described by Lennig may be used to combine multiple prosodic knowledge sources, or to combine one or more prosodic knowledge sources with one or more non-prosodic knowledge sources, to detect an end-of-utterance condition. The technique involves creating a histogram, based on training data, for each knowledge source. Training data consists of both "positive" and "negative" utterances. Positive utterances are defined as those utterances which meet the

criterion of interest (e.g., end-of-utterance), while negative utterances are defined as those utterances which do not. Each knowledge source is represented as a scalar value. The bin boundaries of each histogram partition the range of the feature into a number of bins. These boundaries are determined empirically so that there is enough resolution to distinguish useful differences in values of the knowledge source but so that there is a sufficient amount of data in each bin. The bins need not be of uniform width.

It may be useful to smooth the histograms, particularly when there is limited training data. One approach to doing so is "medians of three" smoothing, described in J.W. Tukey, "Smoothing Sequences," Exploratory Data Analysis, Addison-Wesley, 1977. In medians of three smoothing, starting at one end of the histogram and processing each bin in order until reaching the other end, the count of each bin is replaced by the median of the counts of that bin and the two adjacent bins. The smoothing is applied separately to the positive and negative bin counts.

At run time, a given knowledge source (e.g., intonation) is measured. The value of this knowledge source determines the histogram bin into which it falls. Suppose that bin is bin number K. Let A represent the number of positive training utterances that fell into bin K and let B represent the number of negative training utterances that fell into bin K. A probability score P_i of this knowledge source is then computed as $P_i = A/(A+B)$, where P_i represents the probability that the criterion of interest is satisfied given the current value of this knowledge source. The same process is used for each additional knowledge source. The probabilities of the different knowledge sources are then combined to generate an overall probability P as follows: $P = (P_1{**}w_1)(P_2{**}w_2)(P_3{**}w_3)\ldots(P_N{**}w_N)$, where the "**" operator indicates exponentiation and w_1, w_2, w_3, etc. are empirically-determined, non-negative weights that sum to one.

Intonation of an utterance is one prosodic knowledge source that can be useful in endpoint detection. Various techniques can be used to determine the intonation. The intonation of an utterance is represented, at least in part, by the change in fundamental frequency of the utterance over time. Hence, the intonation of an utterance may be determined in the form of a pattern (an

"intonation pattern") indicating the change in fundamental frequency of the utterance over time. In the English language, a generally decreasing fundamental frequency is more indicative of an end-of-utterance condition than a generally increasing fundamental frequency. Hence, a decline in fundamental frequency may represent decreasing intonation, which may be evidence of an end-of-utterance condition.

There are many possible approaches to mapping a declining fundamental frequency pattern into a scalar feature, for use in the above-described histogram approach. The intonation pattern may be, for example, a single computation based on the difference in fundamental frequency between two frames of data, or it may be based on multiple differences for three or more (potentially overlapping) frames within a predetermined time range. For this purpose, it may be sufficient to examine the most recent approximately 0.6 to 1.2 seconds or one to three syllables of speech.

One specific approach involves computing the smoothed first difference of the fundamental frequency. Let $F(n)$ represent the fundamental frequency, F0, of frame n. Let $F'(n) = F(n) - F(n-1)$ represent the first difference of $F(n)$. Let $f(n) = aF'(n) - (1-a)f(n-1)$, where $0 \leq a \leq 1$, represent the smoothed first difference of $F(n)$. The value of "a" is tuned empirically so that $f(n)$ becomes as negative as possible when the F0 pattern declines at the end of an utterance. Use $f(n)$ as an input feature to the histogram method. Note that when $F(n)$ is undefined because it is in an unvoiced segment of speech, $F(n)$ may be defined as $F(n-1)$.

Other approaches could capture more information about the time evolution of the fundamental frequency pattern using techniques such as Hidden Markov Models, where the parameter $f(n)$ is the observation parameter.

The intonation pattern may additionally (or alternatively) include the relationship between the current fundamental frequency and the fundamental frequency range of the speaker. For example, a drop in fundamental frequency to a value that is near the low end of the fundamental frequency range of the speaker may suggest an end-of-utterance condition. It may be desirable to treat as two distinct knowledge sources the change in fundamental frequency over time and

the relationship between the current fundamental frequency and the speaker's fundamental frequency range. In that case, these two intonation-based knowledge sources may be combined using the above-described histogram approach, for purposes of detecting an end-of-utterance condition.

To apply the histogram approach to the latter-mentioned knowledge source, the low end of the speaker's fundamental frequency range is computed as a scalar. One way of doing this is simply to use the minimum observed fundamental frequency for the speaker. The fundamental frequency range of the speaker may be determined adaptively from utterances of the speaker earlier in a dialog. In one embodiment, the system asks the speaker a question specifically designed to elicit a response conducive to determining the low end of the speaker's fundamental frequency range. This may be a simple yes/no question, the response of which will normally contain the word "yes" or "no" with a falling intonation approaching the low end of the speaker's fundamental frequency range. The fundamental frequency of the vowel of the speaker's response may be used as an initial estimate of the low end of the speaker's fundamental frequency range. However this low end of the fundamental frequency range is estimated, designate it as C. Hence, the value input to the fundamental frequency range histogram may be computed as F0 - C.

Any of various knowledge sources may be used as input in the histogram technique described above, to compute the probability P. These knowledge sources may include, for example, any one or more of the following: silence duration, silence duration normalized for peaking rate, f(n) as defined above, F0 - C as defined above, final syllable duration, final syllable duration normalized for phonemic content, final syllable duration normalized for stress, or final syllable duration normalized for a combination of the foregoing parameters.

Various non-histogram based approaches can also be used to perform prosody based endpoint detection. Figure 4 illustrates a non-histogram based approach for prosody based determination of an end-of-utterance condition, according to one embodiment, which may be implemented in the endpoint detector 5. Initially, the next frame of speech is input to the endpoint detector 5 at

401. It is next determined at 402 whether the log energy (the logarithm of the energy of the speech signal) is below a predetermined energy threshold level. This threshold level may be set dynamically and adaptively. The specific value of the threshold level may also depend on various factors, such as the specific application of the system and desired system performance, and is therefore not provided herein. If the log energy is not below the threshold level, the process repeats from 401. If the log energy is below the threshold level, then at 403 the intonation pattern of the utterance is determined, which may be done as described above.

Next, at 404 the intonation pattern is referenced against an intonation model to determine a preliminary probability P_1 that the end-of the utterance condition has been reached, given that intonation pattern. The intonation model may be one of prosody models 3-1 through 3-N in Figure 1 and may be in the form of a histogram based on training data, such as described above. Other examples of the format of the intonation model are described below. In essence, this is a determination of whether the intonation pattern is suggestive of an end-of-utterance condition. As noted above, a generally decreasing intonation may suggest an end-of-utterance condition. Again, it may be sufficient to examine the last approximately 0.6 to 1.2 seconds or one to three syllables of speech for this purpose.

As noted above, other intonation-based parameters (e.g., the relationship between the fundamental frequency and the speaker's fundamental frequency range) may be represented in the intonation model. Alternatively, such other parameters may be treated as separate knowledge sources and referenced against separate intonation models to obtain separate probability values.

Referring still to Figure 4, at 405 the amount of time T_1 which the speech signal has remained below the energy threshold level is computed. This amount of time T_1 is then referenced at 406 against a model of elapsed time to determine a second preliminary probability P_2 that the end-of-utterance has been reached,

given the pause duration T_1. At 407, the normalized, relative duration T_2 of the final syllable of the utterance is computed. Although the duration of the final syllable of the utterance cannot actually be known before an end-of-utterance condition has been identified, this computation 407 may be based on the temporary assumption (i.e., only for purposes of this computation) that an end-of-utterance condition has occurred. Techniques for automatically determining the duration of a syllable of an utterance are well-known. Once computed, the duration T_2 is then referenced at 408 against a syllable duration model (e.g., another one of prosody models 3-1 through 3-N) to determine a third preliminary probability P_3 of end-of-utterance, given the normalized relative duration T_2 of the last syllable.

At 409, the overall probability P of end-of-utterance is computed as a function of P_1, P_2 and P_3, which may be, for example, a geometrically weighted average of P_1, P_2 and P_3. In this computation, each probability value P_1, P_2 and P_3 is raised to a power, so that the sum of these three probabilities equals one. At 410, the overall probability P is compared against a threshold probability level P_{th}. If P exceeds the threshold probability P_{th} at 410, then an end-of-utterance is determined to have occurred at 411, and the process then repeats from 401. Otherwise, an end-of-utterance is not yet identified, and the process repeats from 401. The threshold probability P_{th} as well as the specific or other function used to compute the overall probability P can depend upon various factors, such as the particular application of the system, the desired performance, etc.

Many variations upon this process are possible, as will be recognized by those skilled in the art. For example, the order of the operations mentioned above may be changed for different embodiments.

Referring again to operation 404 in Figure 4, the intonation model may have any of a variety of possible forms, an example of which is a histogram based on training data. In yet another approach, the intonation model may be a

regression model or a Gaussian distribution of training data, with an estimated mean and variance, against which the input data is compared to assign the probability values P_1. Parametric approaches such as these can optionally be implemented using a Hidden Markov Model to capture information about the time evolution of the intonation pattern.

As an example of a non-parametric approach, the intonation model may be a prototype function of declining fundamental frequency over time (i.e., representing known end-of-utterance conditions). Thus, the operation 404 may be accomplished by computing the correlation between the observed intonation pattern and the prototype function. In this approach, it may be useful to express the prototype function and the observed intonation values as percentage increases or decreases in fundamental frequency, rather than as absolute values.

As yet another example, the intonation model may be a simple look-up table of intonation patterns (i.e., functions or values) vs. probability values P_1. Interpolation may be used to map input values that do not exactly match a value in the table.

Referring to operation 406 in Figure 4, the model of elapsed time (during which the speech has exhibited low energy) may also include a histogram constructed from training data, or another format such as described above. Since different speech recognition grammars may give rise to different post-speech timeout parameters, it may be useful to introduce an additive bias that is adjustable through tuning, to the computation of probability P_2. This additive bias may be subtracted from the observed length of time T_1 of low energy speech before using the result to compute probability P_2 using the histogram approach. This approach would provide the system designer with the ability to bias the system to require longer silences to conclude an end-of-utterance has occurred.

Referring to operation 408 in Figure 4, the syllable duration model may have essentially any form that is suitable for this purpose, such as a histogram or other format described above.

Figures 5A and 5B collectively represent another embodiment of the prosody based endpoint detection technique. The processes of Figures 5A and 5B may be performed concurrently. The process of Figure 5A is for determining a threshold time value T_{th}, which is used in the process of Figure 5B to identify an end-of-utterance condition. Specifically, the threshold time value T_{th} determines how long the endpoint detector will wait, in response to detecting the input signal's log energy has fallen below a threshold level, before determining an end-of-utterance has occurred.

Referring first to Figure 5A, initially the next frame of speech representing an utterance is input at 501. At 502, the intonation pattern of the utterance is determined, such as in the manner described above. At 503, a determination is made of whether the intonation pattern is generally suggestive of (e.g., in terms of probability) an end-of-utterance condition. This determination 503 may be made in the manner described above. If the intonation of the utterance is determined at 503 to be suggestive of an end-of-utterance condition, then at 505 the threshold time value T_{th} is set equal to a predetermined time value y. If not, then at 504 the threshold time value T_{th} is set equal to a predetermined time value x, which is larger than (represents longer duration than) time value y. The specific values for x and y can depend upon various factors, such as the particular application of the system, the desired performance, etc.

Referring now to Figure 5B, a timer variable T_4 is initialized to zero at 510, and at 511 the next frame of speech is input. At 512, a determination is made of whether the log energy of the speech has dropped below the threshold level. If not, T_4 is reset to zero at 516, and the process then repeats from 511. If the signal has dropped below the threshold level, then at 513 T_4 is incremented. Next, at 514 T_4 is compared to the threshold time value T_{th} determined in the process of Figure 5A. If T_4 exceeds T_{th}, then at 515 an end-of-utterance condition is identified, and the process repeats from 510. Otherwise, an end-of-utterance condition is not yet

identified, and the process repeats from 511. Many variations upon these processes are possible without altering the basic approach, such as changing the ordering of the above-noted operations.

Thus, a method and apparatus for detecting endpoints of speech using prosody have been described. Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention as set forth in the claims. Accordingly, the specification and drawings are to be regarded in an illustrative sense rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A method comprising:

inputting speech representing an utterance and having an intonation; and

identifying an endpoint of the utterance based on the intonation.

2. A method as recited in claim 1, wherein said identifying an endpoint of the utterance based on the intonation comprises comparing the intonation with an intonation model.

3. A method as recited in claim 4, further comprising determining the intonation by computing the fundamental frequency of the utterance.

4. A method as recited in claim 3, wherein said determining the intonation comprises using an intonation model to determine the intonation.

5. A method as recited in claim 1, wherein said identifying the endpoint of the utterance comprises identifying the endpoint of the utterance based on a plurality of knowledge sources, wherein one of the knowledge sources is intonation, including referencing the input speech against a histogram based on training data for each of the knowledge sources.

6. A method as recited in claim 1, further comprising:

determining a period of time that has elapsed since the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

7. A method as recited in claim 1, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on a

length of time for which an energy value of the speech has remained below a predetermined energy value.

8. A method as recited in claim 7, wherein said identifying an endpoint of the utterance further comprises identifying the endpoint of the utterance based on the duration of the final syllable of the utterance.

9. A method of operating an endpoint detector, the method comprising:

inputting speech representing an utterance, the utterance having an intonation; and

comparing the intonation of the utterance with an intonation model;

determining a probability based on a result of said comparing; and

identifying an endpoint of the utterance based on the probability.

10. A method as recited in claim 9, further comprising determining the intonation of the utterance as a function of the fundamental frequency of the utterance.

11. A method as recited in claim 9, further comprising:

determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

12. A method as recited in claim 9, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the duration of the final syllable of the utterance.

13. A method as recited in claim 12, wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on a

period of time for which an energy value of the speech has remained below a threshold value.

14. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

determining that a value of the speech has dropped below a threshold value;

computing an intonation of the utterance;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a period of time that has elapsed since the value of the speech dropped below the threshold value;

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first and second end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred based on the overall end-of-utterance probability.

15. A method as recited in claim 14, wherein said computing an intonation of the utterance comprises computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time.

16. A method as recited in claim 15, further comprising:

determining a duration of a final syllable of the utterance; and

referencing the duration of the final syllable against a syllable duration model to determine a third end-of-utterance probability;

wherein said computing an overall end-of-utterance probability comprises computing the overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities.

17. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

computing an intonation of the utterance;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a duration of a final syllable of the utterance;

referencing the duration of the final syllable against a syllable duration model to determine a second end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first and second end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred based on the overall end-of-utterance probability.

18. A method as recited in claim 17, wherein said computing an intonation of the utterance comprises computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time.

19. A method as recited in claim 17, further comprising:

determining that a value of the speech has dropped below a threshold value;

determining a period of time that has elapsed since the value of the speech dropped below the threshold value; and

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

wherein said computing an overall end-of-utterance probability comprises computing the overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities.

20. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

determining that a value of the speech has dropped below a threshold value;

computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time;

referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

determining a period of time that has elapsed since a value of the speech dropped below the threshold value;

referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

determining a duration of a final syllable of the utterance;

referencing the duration of the final syllable against a syllable duration model to determine a third end-of-utterance probability;

computing an overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities; and

determining whether an end-of-utterance has occurred by comparing the overall end-of-utterance probability to a threshold probability.

21. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance;

determining an intonation of the utterance;

if the intonation of the utterance is determined to be generally decreasing, then setting a threshold time period equal to a first time value;

if the intonation of the utterance is determined not to be generally decreasing, then setting the threshold time period equal to a second time value larger than the first time value; and

identifying an endpoint of the utterance based on the threshold time period.

22. A method as recited in claim 21, wherein said using the threshold time period to identify an endpoint of the utterance comprises using the threshold time period to identify an endpoint of the utterance by determining that an endpoint of the utterance has occurred if an energy value of the speech remains below a predetermined value for the threshold time period.

23. A method as recited in claim 21, wherein said determining an intonation of the utterance comprises using an intonation model.

24. A method of operating an endpoint detector for speech recognition, the method comprising:

inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

determining an intonation of the utterance by

computing the intonation as the fundamental frequency of the utterance as a function of time, and

referencing the intonation against an intonation model to determine the intonation of the utterance;

if the intonation of the utterance is determined to be generally decreasing, then setting a threshold time period equal to a first time value;

if the intonation of the utterance is determined not to be generally decreasing, then setting the threshold time period equal to a second time value larger than the first time value; and

using the threshold time period to identify an endpoint of the utterance, by determining that an endpoint of the utterance has occurred if the speech remains below a predetermined value for a length of time equal to the threshold time period.

25. A machine-readable program storage medium tangibly embodying a sequence of instructions executable by a machine to perform a method for endpoint detection, the method comprising:

inputting speech representing an utterance, the utterance having an intonation; and

identifying an endpoint of the utterance based on the intonation of the utterance.

26. A machine-readable program storage medium as recited in claim 25, wherein said using the intonation of the utterance in identifying an endpoint of the utterance comprises comparing the intonation of the utterance with an intonation model.

27. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises determining the intonation of the utterance.

28. A machine-readable program storage medium as recited in claim 27, wherein said determining the intonation of the utterance comprises computing the fundamental frequency of the utterance.

29. A machine-readable program storage medium as recited in claim 27, wherein said determining the intonation of the utterance comprises using an intonation model to determine the intonation of the utterance.

30. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises:

determining a period of time for which an energy value of the speech has been below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

31. A machine-readable program storage medium as recited in claim 25, wherein the method further comprises:

determining a duration of a final syllable of the utterance; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the duration of a final syllable of the utterance.

32. A machine-readable program storage medium as recited in claim 31, wherein the method further comprises:

determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said identifying an endpoint of the utterance comprises identifying the endpoint of the utterance further based on the period of time.

33. An endpoint detector comprising:

means for inputting speech representing an utterance, the utterance having an intonation; and

means for identifying an endpoint of the utterance based on the intonation of the utterance.

34. An endpoint detector as recited in claim 33, wherein said means for using the intonation of the utterance in identifying an endpoint of the utterance comprises means for comparing the intonation of the utterance with an intonation model.

35. An endpoint detector as recited in claim 33, further comprising means for determining the intonation of the utterance.

36. An endpoint detector as recited in claim 35, wherein said means for determining the intonation of the utterance comprises means for computing the fundamental frequency of the utterance.

37. An endpoint detector as recited in claim 35, wherein said means for determining the intonation of the utterance comprises means for using an intonation model to determine the intonation of the utterance.

38. An endpoint detector as recited in claim 33, further comprising:

means for determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the period of time.

39. An endpoint detector as recited in claim 33, further comprising:

means for determining a duration of a final syllable of the utterance; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the duration of a final syllable of the utterance.

40. An endpoint detector as recited in claim 39, further comprising:

means for determining a period of time that has elapsed since a value of the speech dropped below a threshold value; and

wherein said means for identifying an endpoint of the utterance comprises means for identifying the endpoint of the utterance further based on the period of time.

41. An apparatus for performing endpoint detection comprising:

means for inputting speech representing an utterance, the utterance having a time-varying fundamental frequency;

means for determining that a value of the speech has dropped below a threshold value;

means for computing an intonation of the utterance by determining the fundamental frequency of the utterance as a function of time;

means for referencing the intonation of the utterance against an intonation model to determine a first end-of-utterance probability;

means for determining a period of time that has elapsed since the speech dropped below the threshold value;

means for referencing the period of time against an elapsed time model to determine a second end-of-utterance probability;

means for referencing the duration of the final syllable of the utterance against a syllable duration model to determine a third end-of-utterance probability;

means for computing an overall end-of-utterance probability as a function of the first, second, and third end-of-utterance probabilities; and

means for determining whether an end-of-utterance has occurred by comparing the overall end-of-utterance probability to a threshold probability.

ABSTRACT OF THE DISCLOSURE

A method and apparatus are provided for performing prosody based endpoint detection of speech in a speech recognition system. Input speech represents an utterance, which has an intonation pattern. An end-of-utterance condition is identified based on prosodic parameters of the utterance, such as the intonation pattern and the duration of the final syllable of the utterance, as well as non-prosodic parameters, such as the log energy of the speech.

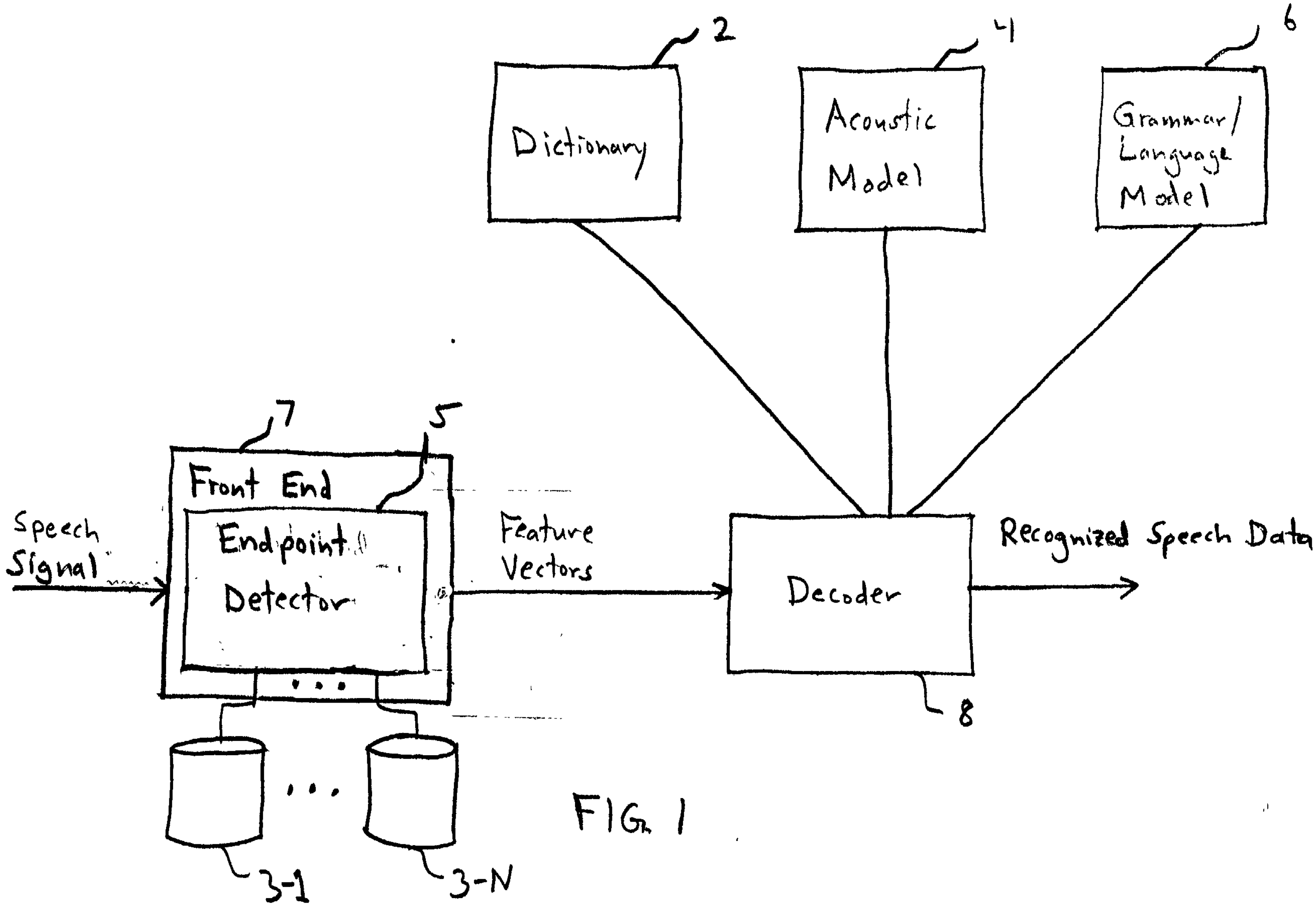

Dictionary
2
Acoustic Model
4
Grammar/ Language Model
6
Front End
7
Endpoint Detector
5
Speech Signal
Feature Vectors
Decoder
8
Recognized Speech Data
3-1
3-N

FIG. 1

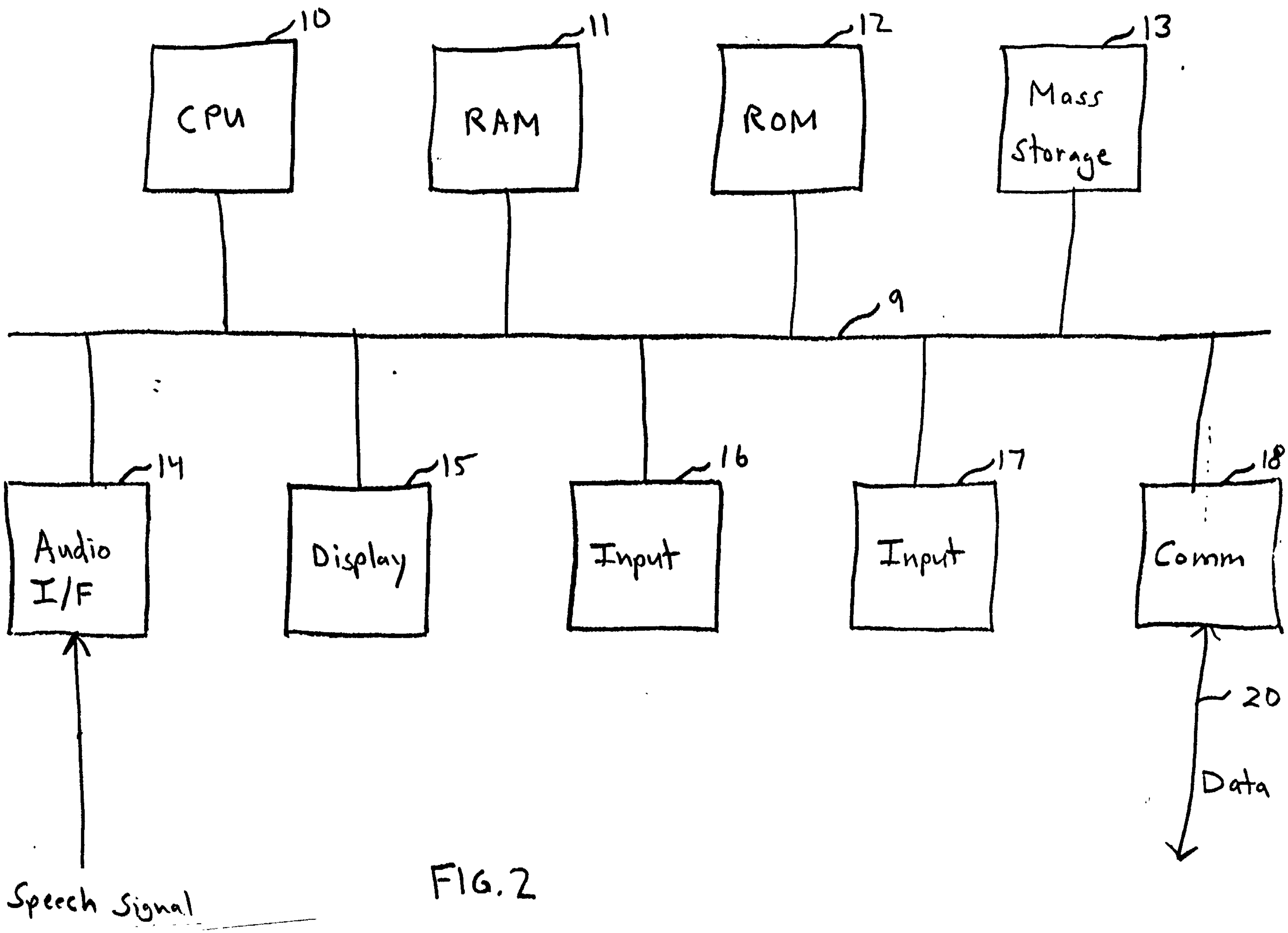

10
CPU
11
RAM
12
ROM
13
Mass Storage
9
14
Audio I/F
15
Display
16
Input
17
Input
18
Comm
20
Data
Speech Signal

FIG. 2

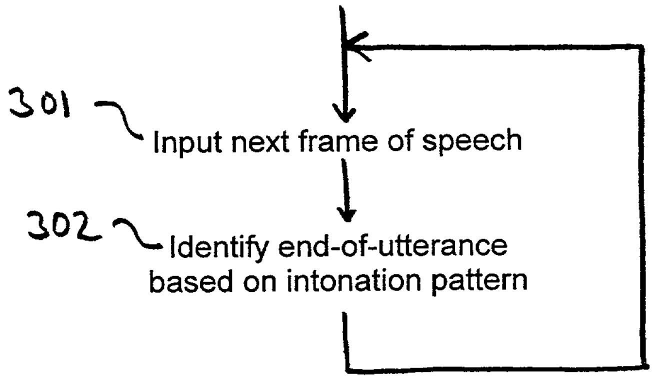
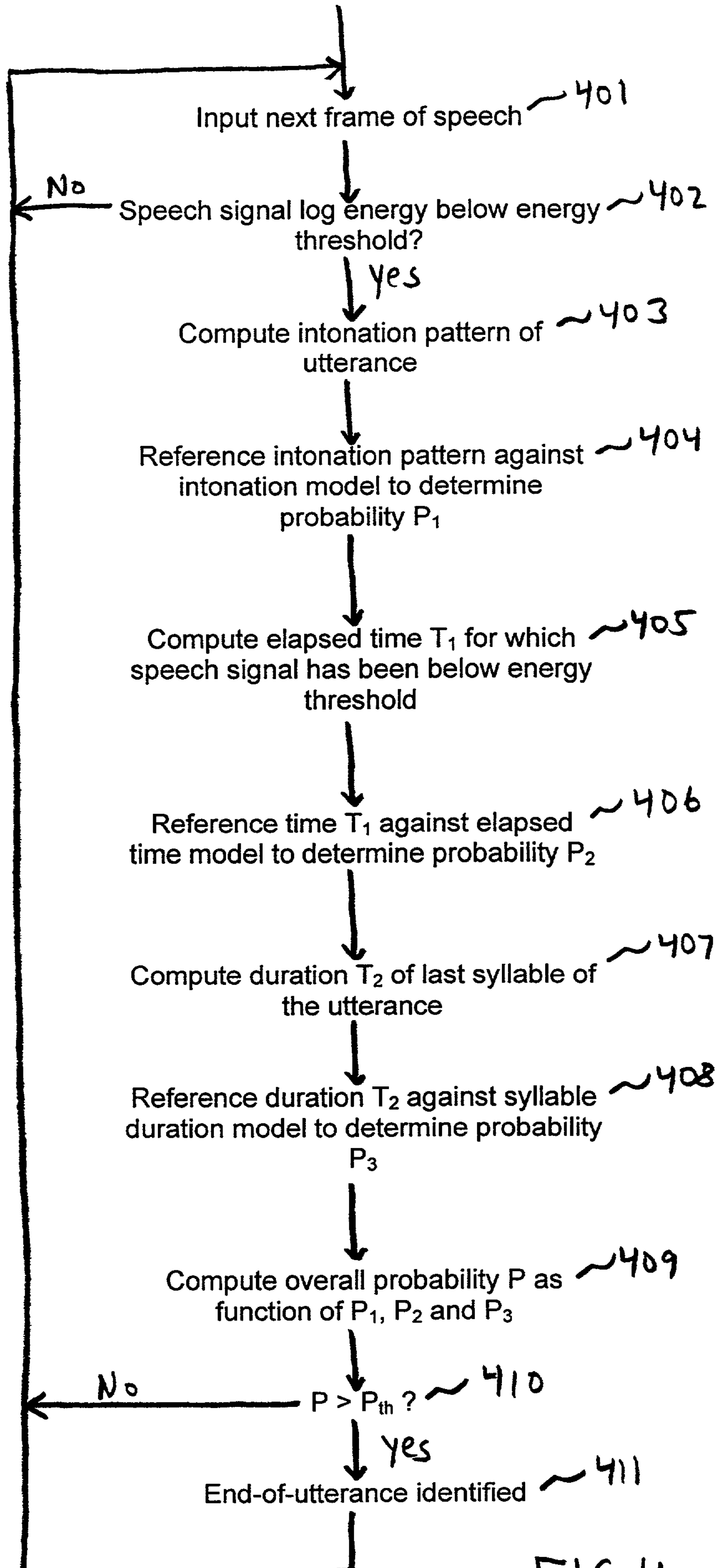
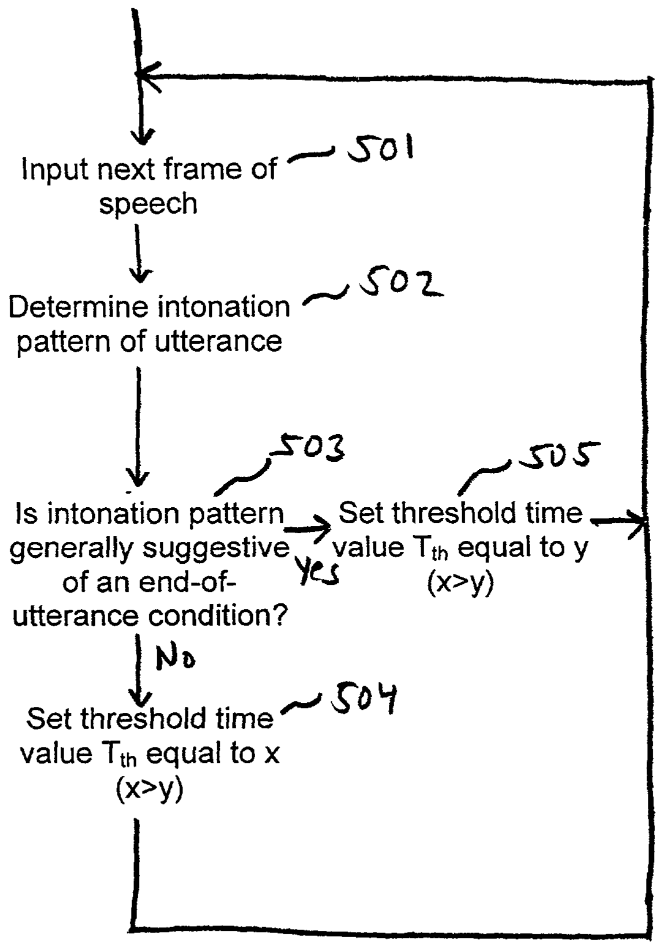

301
Input next frame of speech
302
Identify end-of-utterance based on intonation pattern

FIG.3



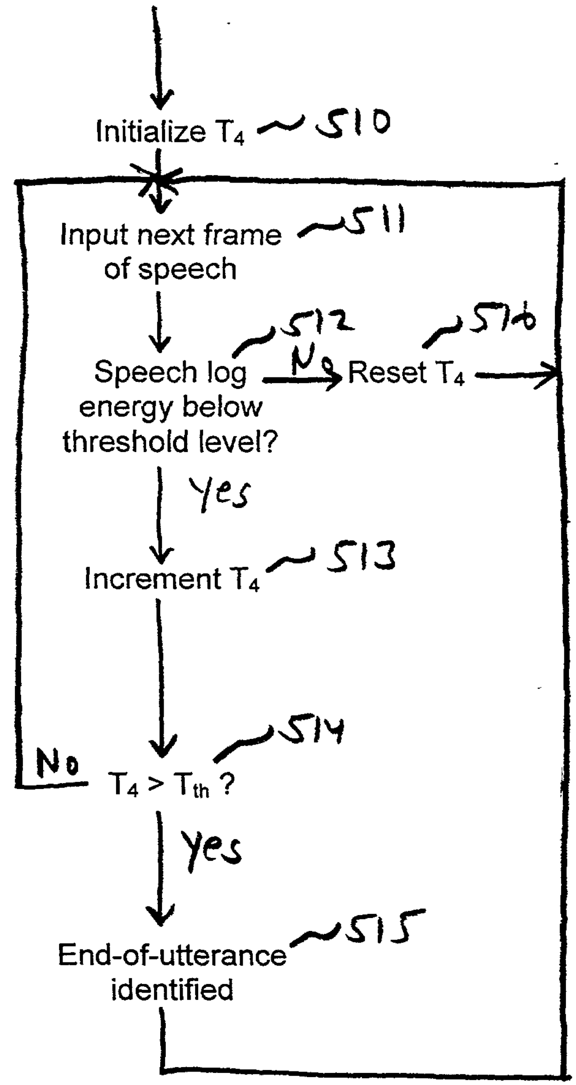
Input next frame of speech
401
No
Speech signal log energy below energy threshold?
402
yes
Compute intonation pattern of utterance
403
Reference intonation pattern against intonation model to determine probability P_1
404
Compute elapsed time T_1 for which speech signal has been below energy threshold
405
Reference time T_1 against elapsed time model to determine probability P_2
406
Compute duration T_2 of last syllable of the utterance
407
Reference duration T_2 against syllable duration model to determine probability P_3
408
Compute overall probability P as function of P_1, P_2 and P_3
409
No
$P > P_{th}$?
410
yes
End-of-utterance identified
411

FIG. 4



501
Input next frame of speech
502
Determine intonation pattern of utterance
503
Is intonation pattern generally suggestive of an end-of-utterance condition?
Yes
505
Set threshold time value T_{th} equal to y (x>y)
No
504
Set threshold time value T_{th} equal to x (x>y)

FIG. 5A



Initialize T_4
510
511
Input next frame of speech
512
516
Speech log energy below threshold level?
No
Reset T_4
Yes
Increment T_4
513
514
No
$T_4 > T_{th}$?
Yes
End-of-utterance identified
515

FIG. 5B

Attorney's Docket No.: 003932.P014 Patent

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below, next to my name.

I believe I am the original, first, and sole inventor (if only one name is listed below) or an original, first, and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled
Prosody Based Endpoint Detection

the specification of which

X is attached hereto.
___ was filed on ______________________ as
United States Application Number ______________________
or PCT International Application Number ______________________
and was amended on ______________________.
(if applicable)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claim(s), as amended by any amendment referred to above. I do not know and do not believe that the claimed invention was ever known or used in the United States of America before my invention thereof, or patented or described in any printed publication in any country before my invention thereof or more than one year prior to this application, that the same was not in public use or on sale in the United States of America more than one year prior to this application, and that the invention has not been patented or made the subject of an inventor's certificate issued before the date of this application in any country foreign to the United States of America on an application filed by me or my legal representatives or assigns more than twelve months (for a utility patent application) or six months (for a design patent application) prior to this application.

I acknowledge the duty to disclose all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d), of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed	
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No

I hereby claim the benefit under title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below:

(Application Number)	Filing Date
(Application Number)	Filing Date

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Number)	Filing Date	(Status -- patented, pending, abandoned)
(Application Number)	Filing Date	(Status -- patented, pending, abandoned)

I hereby appoint the persons listed on Appendix A hereto (which is incorporated by reference and a part of this document) as my respective patent attorneys and patent agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected herewith.

Send correspondence to Jordan M. Becker, BLAKELY, SOKOLOFF, TAYLOR &
(Name of Attorney or Agent)
ZAFMAN LLP, 12400 Wilshire Boulevard 7th Floor, Los Angeles, California 90025 and direct telephone calls to Jordan M. Becker, (408) 720-8300.
(Name of Attorney or Agent)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of Sole/First Inventor Matthew Lennig

Inventor's Signature ______________________ Date ______________

Residence Palo Alto, California Citizenship U.S.A.
(City, State) (Country)

Post Office Address 1380 Willow Road
Menlo Park, California 94025

Full Name of Second/Joint Inventor ______________________

Inventor's Signature ______________________ Date ______________

Residence ______________________ Citizenship ______________
(City, State) (Country)

Post Office Address ______________________

Full Name of Third/Joint Inventor ______________________

Inventor's Signature ______________________ Date ______________

Residence ______________________ Citizenship ______________
(City, State) (Country)

Post Office Address ______________________

Full Name of Fourth/Joint Inventor ______________________

Inventor's Signature ______________________ Date ______________

Residence ______________________ Citizenship ______________
(City, State) (Country)

Post Office Address ______________________

APPENDIX A

William E. Alford, Reg. No. 37,764; Farzad E. Amini, Reg. No. P42,261; Aloysius T. C. AuYeung, Reg. No. 35,432; William Thomas Babbitt, Reg. No. 39,591; Carol F. Barry, Reg. No. 41,600; Jordan Michael Becker, Reg. No. 39,602; Bradley J. Bereznak, Reg. No. 33,474; Michael A. Bernadicou, Reg. No. 35,934; Roger W. Blakely, Jr., Reg. No. 25,831; Gregory D. Caldwell, Reg. No. 39,926; Ronald C. Card, Reg. No. 44,587; Andrew C. Chen, Reg. No. 43,544; Thomas M. Coester, Reg. No. 39,637; Alin Corie, Reg. No. P46,244; Dennis M. deGuzman, Reg. No. 41,702; Stephen M. De Klerk, under 37 C.F.R. § 10.9(b); Michael Anthony DeSanctis, Reg. No. 39,957; Daniel M. De Vos, Reg. No. 37,813; Robert Andrew Diehl, Reg. No. 40,992; Sanjeet Dutta, Reg. No. P46,145; Matthew C. Fagan, Reg. No. 37,542; Tarek N. Fahmi, Reg. No. 41,402; Paramita Ghosh, Reg. No. 42,806; James Y. Go, Reg. No. 40,621; James A. Henry, Reg. No. 41,064; Willmore F. Holbrow III, Reg. No. P41,845; Sheryl Sue Holloway, Reg. No. 37,850; George W Hoover II, Reg. No. 32,992; Eric S. Hyman, Reg. No. 30,139; William W. Kidd, Reg. No. 31,772; Sang Hui Kim, Reg. No. 40,450; Eric T. King, Reg. No. 44,188; Erica W. Kuo, Reg. No. 42,775; Kurt P. Leyendecker, Reg. No. 42,799; Michael J. Mallie, Reg. No. 36,591; Andre L. Marais, under 37 C.F.R. § 10.9(b); Paul A. Mendonsa, Reg. No. 42,879; Darren J. Milliken, Reg. 42,004; Lisa A. Norris, Reg. No. 44,976; Chun M. Ng, Reg. No. 36,878; Thien T. Nguyen, Reg. No. 43,835; Thinh V. Nguyen, Reg. No. 42,034; Dennis A. Nicholls, Reg. No. 42,036; Daniel E. Ovanezian, Reg. No. 41,236; Marina Portnova, Reg. No. P45,750; Babak Redjaian, Reg. No. 42,096; William F. Ryann, Reg. 44,313; James H. Salter, Reg. No. 35,668; William W. Schaal, Reg. No. 39,018; James C. Scheller, Reg. No. 31,195; Jeffrey Sam Smith, Reg. No. 39,377; Maria McCormack Sobrino, Reg. No. 31,639; Stanley W. Sokoloff, Reg. No. 25,128; Judith A. Szepesi, Reg. No. 39,393; Vincent P. Tassinari, Reg. No. 42,179; Edwin H. Taylor, Reg. No. 25,129; John F. Travis, Reg. No. 43,203; George G. C. Tseng, Reg. No. 41,355; Joseph A. Twarowski, Reg. No. 42,191; Lester J. Vincent, Reg. No. 31,460; Glenn E. Von Tersch, Reg. No. 41,364; John Patrick Ward, Reg. No. 40,216; Mark L. Watson, Reg. No. P46,322; Thomas C. Webster, Reg. No. P46,154; Charles T. J. Weigell, Reg. No. 43,398; Kirk D. Williams, Reg. No. 42,229; James M. Wu, Reg. No. 45,241; Steven D. Yates, Reg. No. 42,242; and Norman Zafman, Reg. No. 26,250; my patent attorneys, and Justin M. Dillon, Reg. No. 42,486; my patent agent, of BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP, with offices located at 12400 Wilshire Boulevard, 7th Floor, Los Angeles, California 90025, telephone (310) 207-3800, and James R. Thein, Reg. No. 31,710, my patent attorney.

APPENDIX B

Title 37, Code of Federal Regulations, Section 1.56
Duty to Disclose Information Material to Patentability

(a) A patent by its very nature is affected with a public interest. The public interest is best served, and the most effective patent examination occurs when, at the time an application is being examined, the Office is aware of and evaluates the teachings of all information material to patentability. Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability as defined in this section. The duty to disclosure information exists with respect to each pending claim until the claim is cancelled or withdrawn from consideration, or the application becomes abandoned. Information material to the patentability of a claim that is cancelled or withdrawn from consideration need not be submitted if the information is not material to the patentability of any claim remaining under consideration in the application. There is no duty to submit information which is not material to the patentability of any existing claim. The duty to disclosure all information known to be material to patentability is deemed to be satisfied if all information known to be material to patentability of any claim issued in a patent was cited by the Office or submitted to the Office in the manner prescribed by §§1.97(b)-(d) and 1.98. However, no patent will be granted on an application in connection with which fraud on the Office was practiced or attempted or the duty of disclosure was violated through bad faith or intentional misconduct. The Office encourages applicants to carefully examine:

(1) Prior art cited in search reports of a foreign patent office in a counterpart application, and

(2) The closest information over which individuals associated with the filing or prosecution of a patent application believe any pending claim patentably defines, to make sure that any material information contained therein is disclosed to the Office.

(b) Under this section, information is material to patentability when it is not cumulative to information already of record or being made or record in the application, and

(1) It establishes, by itself or in combination with other information, a prima facie case of unpatentability of a claim; or

(2) It refutes, or is inconsistent with, a position the applicant takes in:

(i) Opposing an argument of unpatentability relied on by the Office, or

(ii) Asserting an argument of patentability.

A prima facie case of unpatentability is established when the information compels a conclusion that a claim is unpatentable under the preponderance of evidence, burden-of-proof standard, giving each term in the claim its broadest reasonable construction consistent with the specification, and before any consideration is given to evidence which may be submitted in an attempt to establish a contrary conclusion of patentability.

(c) Individuals associated with the filing or prosecution of a patent application within the meaning of this section are:

(1) Each inventor named in the application;

(2) Each attorney or agent who prepares or prosecutes the application; and

(3) Every other person who is substantively involved in the preparation or prosecution of the application and who is associated with the inventor, with the assignee or with anyone to whom there is an obligation to assign the application.

(d) Individuals other than the attorney, agent or inventor may comply with this section by disclosing information to the attorney, agent, or inventor.